Obesity Levels of Healthy Cities

Kimberly Cable

Belleview University

DSC 530: Data Exploration and Analysis

Dr. Shankar Parajulee

June 4, 2022

Obesity Levels of Healthy Cities?

A study was done gathering data on all the cities in the world to see which were the top healthiest cities. But what makes these cities healthy? In my analysis, I wanted to know what were the determining factors that made them the top healthiest cities in the world. Specifically looking at obesity levels, the number of hours of sunshine a city gets per day, the cost of a bottle of water in the city’s country, the number of outdoor activities a city has, the average number of hours a person works, pollution levels, and life expectancy.

I did some preliminary analysis to find out all the outliers in these variables and determine their descriptive statistics like mean, median, mode, and quartiles. I also did a cumulative distribution function plot of each variable.

Nothing looked too out of the ordinary, so I wanted to know if the number of hours of sunshine affected obesity levels in healthy cities. I determined the most frequent number of hours of sunshine per day was five, so I divided the data into two groups. First those cities with five hours of sunshine and the rest of the cities. With this, I looked at the obesity levels for each group. I found that with just five hours of sunshine the obesity levels were much lower than in the rest of the cities not having an average of five hours of sunshine per day.

Checking if the obesity levels were of a normal distribution, I did a probability plot and found that the ends were skewed, and the middle fell just short of being a normal distribution.

Next, I compared obesity levels with the number of hours of sunshine per day with a scatterplot, this suggested that there is a positive correlation between the number of hours of sunshine and obesity levels. The Pearson’s correlation coefficient of 0.31 with a p-value of 0.001 and an R-squared of 0.09 confirms that with a coefficient and significant p-value there is no correlation between the number of hours of sunshine per day and obesity levels in healthy cities. I then looked to see if there was a correlation between the cost of a bottle of water and the number of outdoor activities a healthy city has. The scatter plot indicates as the cost of a bottle of water goes up the number of outdoor activities goes down. The Pearson’s coefficient is -0.29 with a p-value of 0.056 and an R-squared of 0.086. This also indicates no correlation between the cost of a bottle of water and the number of outdoor activities a city has.

Finally, I did some regression models on how the individual variables influenced obesity levels and then a multiple regression with all of them to see if together they influenced obesity levels in a healthy city. I found that individually, they did not influence the obesity levels but together they did influence the obesity levels but not a significant amount.

Looking at the analysis I performed, I feel I didn’t get very far in determining what obesity levels were in healthy cities. I think having data from non-healthy cities may have helped in the analysis. One assumption I made was that healthy cities had lower obesity levels which in my mind made them healthy which skewed my thinking and choosing the variables from the dataset that I did. With more research and better data, I think determining the obesity levels of healthy cities may come from much more variables and data.